Comparative Genomic Hybridization Analysis of 38 Breast Cancer Cell Lines: A Basis for Interpreting Complementary DNA Microarray Data

Farahnaz Forozan, Eija H. Mahlamäki, Outi Monni, Yidong Chen, Robin Veldman, Yuan Jiang, Gerald C. Gooden, Stephen P. Ethier, Anne Kallioniemi, and Olli-P. Kallioniemi

ABSTRACT

Breast cancer cell lines provide a useful starting point for the discovery and functional analysis of genes involved in breast cancer. Here, we studied 38 established breast cancer cell lines by comparative genomic hybridization (CGH) to determine recurrent genetic alterations and the extent to which these cell lines resemble uncultured tumors. The following chromosomal gains were observed: 8q (75%), 1q (61%), 20q (55%), 7p (44%), 3q (39%), 5p (39%), 7q (39%), 17q (33%), 1p (30%), and 20p (30%), and the most common losses were: 8p (58%), 18q (58%), 1p (42%), Xp (42%), Xq (42%), 4p (36%), 11q (36%), 18p (33%), 10q (30%), and 19p (28%). Furthermore, 35 recurrent high-level amplification sites were identified, most often involving 8q23 (37%), 20q13 (29%), 3q25-q26 (24%), 17q22-q23 (16%), 17q22-24 (16%), 1p13 (11%), 1q22 (11%), 5p13 (11%), 5p14 (11%), 11q13 (11%), 17q12-21 (11%), and 7q12-q22 (11%). A comparison of DNA copy number changes found in the cell lines with those reported in 17 published studies (698 tumors) of uncultured tumors revealed a substantial degree of overlap. CGH copy number profiles may facilitate identification of important new genes located at the hotspots of such chromosomal alterations. This was illustrated by analyzing expression levels of 1236 genes using cDNA microarrays in four of the cell lines. Several highly overexpressed genes (such as RCH1 at 17q23, TOPO II at 17q21-q22, as well as CAS and MYBL2 at 20q13) were involved in these recurrent DNA amplifications.

In conclusion, DNA copy number profiles were generated by CGH for most of the publicly available breast cancer cell lines and were made available on a web site (http://www.nhgri.nih.gov/DIR/CGB/CR2000). This should facilitate the correlation analysis of gene expression and copy number as illustrated here by the finding of cDNA microarrays of several overexpressed genes that were amplified.

INTRODUCTION

Several hundred studies on genetic alterations in cancer have been published using CGH (1). There are at least 19 published CGH studies on breast cancer, reporting genetic alterations in a total of 792 cases. These studies have also identified associations of these genetic alterations with tumor grade (2), histological subtype (3–5), metastasis (6), genetic predisposition (7), as well as patient survival (8). A number of recurrent genetic alterations have been reported, such as gains of 1q, 8q, and 20q, as well as losses of 1p, 16q, and 18q. Furthermore, specific high-level amplifications have been observed at chromosomal sites that do not coincide with the locations of the classical breast cancer oncogenes, such as 1q32, 8q23, 17q23-25q, and 20q11-q13. Several genes have already been implicated to be involved in these amplification sites, including the ZNF217, NABC1, BTAK/STK1/Aurora II, AIB1, and CAS-1 genes at 20q13 (9–12), as well as the ribosomal protein S6 kinase at 17q23 (13, 14). Despite this progress, there are numerous sites in the genome that are often amplified in breast cancer where the genes involved have not yet been identified.

Most of the published studies of breast cancer by CGH have analyzed unique clinical series of breast carcinomas. Therefore, any genetic alterations discovered by CGH can only be followed in the same tissue material, which usually is not accessible to other investigators. Whereas CGH can be easily performed from even small, formalin-fixed, archival tumor tissues, cloning genes involved in genetic rearrangements requires the availability of large quantities of fresh tissue material. Here, we performed a CGH study of 38 breast cancer cell lines, which comprise the majority of the publicly available breast cancer cell lines. Eleven of these represent novel cell lines recently characterized by us (15). The CGH profiles from all 38 cell lines are available on a web site. Furthermore, a cDNA microarray analysis of 1236 genes was performed in four cases to study the relationship of gene expression and genomic copy number.

MATERIALS AND METHODS

CGH. The following breast carcinoma cell lines were studied: MCF7, MDA-436, MDA-361, MDA-468, MDA-157, MDA-175, MDA-453, MDA-134, MDA-231, MDA-435, MDA-415, HS578T, SKBR3, BT474, BT549, BT20, BT483, ZR-75–30, UACC-812, and HBL-100 from American Type Culture Collection (Manassas, VA); ZR-75–1 wt strain from Dr. Jeff Moscov (National Cancer Institute, Bethesda, MD); SUM-149, SUM-1315, SUM-159, SUM-185, SUM-190, SUM-206, SUM-225, and SUM-229, SUM-44, SUM-52, and SUM-102 from Dr. Steve Ethier’s laboratory (University of Michigan); MX-1, KPL-1, CAL-51, COLO-824, EFM-19, and EFM-192A from German Collection of Microorganisms and Cell Cultures (Braunschweig, Germany); and MPE-600 from Vysis Inc. (Downers Grove, IL). CGH was carried out essentially as described previously (15). Normal male versus female hybridizations were used as negative controls and for ensuring the linearity of the hybridization (16). Chromosomal regions where CGH ratios were >1.2 were considered as gained, and those regions where the ratio was <0.8 were considered as lost. High-level amplifications were defined as small regions with a ratio >1.4. To define recurrent, independent regions of amplification, a high-level copy number increase had to be present in a localized chromosomal region in two or more cases.

FISH. FISH with probes to Cyclin-D1, ERBB2, and MYC (Vysis Inc.) was done on 37 of 38 cell lines and with bacterial artificial chromosome probes specific to CAS, RCH1, TOPO II, and MYBL2 on four cell lines. Dual color FISH analysis was done on interphase nuclei as described before (15). The slides were hybridized with SpectrumOrange-labeled probes for the genes of interest with the corresponding SpectrumGreen-labeled centromere probe for chromosomes 8, 11, 17, and 20 as reference probes. Because the cell lines are genetically rather homogeneous, ~20 nonoverlapping nuclei with intact morphology based on 4',6-diamidino-2-phenylindole counterstaining were scored to determine the mean copy number of the gene-specific probes relative to the chromosome-specific reference probe.
Fluorescent cDNA Microarrays. The cDNA microarray used for our experiments consisted of 1236 cDNA clones selected from the Unigene set (17) and printed on glass slides as described earlier (18, 19). Four cell lines with distinct, high-level amplifications were used in the cDNA microarray hybridizations. In all experiments, cancer cell line RNA was compared with normal, mammary gland RNA (Clontech, Palo Alto, CA) as a reference. Total RNA was extracted from breast cancer cell lines using the RNeasy kit (Qiagen, Valencia, CA). The labeling, hybridization, and washing procedures were done as described previously (20). Briefly, 60–100 μg of total RNA extracted from tumor cell lines, and 100 μg of total RNA obtained from control sample were reverse-transcribed by oligo dT-primed polymerization using Superscript II reverse transcriptase (LTI Inc.) and Cy5-dUTP and Cy3-dUTP as fluorescent nucleotides (Amersham). After the labeling, the Cy5- and Cy3-labeled cDNAs were combined with 8 μg of poly(dA) (Pharmacia, Bridgewater, NJ), 4 μg of Escherichia coli tRNA (Sigma Chemical Co., St. Louis, MO), and 10 μg of Cot-1 DNA (Life Technologies, Inc., Rockville, MD) in 0.15% SDS, and 3× SSC. The hybridization mixture was incubated for 2 min at 98°C and then for 10 s at 4°C, and it was applied to a cDNA microarray slide. The hybridization was carried out for 16 h at 65°C. The slides were washed in 0.5× SSC and 0.01% SDS for 2 min each at room temperature.

Fluorescence intensities of the 1236 probe targets on the slide were measured using a custom-designed laser confocal microscope containing a scanning stage, appropriate excitation and emission filters, and two photomultiplier tube detectors for two fluorescent emission channels (20). Intensity data were integrated over 225-μm² pixels and recorded at 16 bits. The Cy3 and Cy5 images were scanned independently through two separated channels, the pixel intensities of which were integrated over a 15-μm² area and recorded at 16 bits. The color images were formed by assigning tumor intensity values into the red channel and control intensity into the green channel.

The data obtained were analyzed using the ArraySuite program, developed at the National Human Genome Research Institute on the IPLab Spectrum platform (21). To determine the actual target region based on the information from both red and green pixel values, a segmentation method was used. After background subtraction, fluorescent intensity of a particular target on the slide was calculated by averaging the intensity of every pixel inside the detected target region, and the difference of expression levels of the target was determined by taking the ratios of the R:G. The ratios were normalized for differential efficiencies of labeling, hybridization, and detection based on 88 preselected internal control genes that are usually stable for most experiments (R:G ratio close to 1.0; Ref. 20). A 99% confidence interval was used throughout the experiments to test the significance of differentially expressed genes. A gene is determined up- or down-regulated when the tumor:normal ratios of expression are not within the 99% confidence interval. The statistical significance of the data for each experiment is determined within that experiment and not across experiments. For this analysis, we used a filter that included all genes exhibiting an intensity at least ±1000 (on a scale of 0–65535 fluorescent units) for either the red (Cy5) or green (Cy3) channels.

RESULTS

CGH Studies on Chromosomal Gains, Losses, and Amplifications. On average, 19 genetic changes were found in breast cancer cell lines (range from 14 to 60), including 9 losses (range from 4 to 25) and 10 gains (range from 3 to 42)/cell line. An example of CGH ratio profiles is shown in Fig. 1, and a summary of the different regions of gains and losses is shown in an ideogram format in Fig. 2. The most common gains were at: 8q (75%), 1q (61%), 20q (55%), 7p (44%), 3q (39%), 5p (39%), 7q (39%), 17q (33%), 1p (30%), and 20p (30%). In addition, partial or complete losses of chromosomal regions were observed at: 8p (58%), 18q (58%), 1p (42%), Xp (42%), Xq (42%), 4p (36%), 11q (36%), 18p (33%), 10q (30%), and 19p (28%).

The CGH results from the cell lines were compared with previously published CGH results from uncultured tumors (Fig. 3). Seventeen published studies, together reporting CGH results from 698 tumors, were analyzed (2, 3, 5, 6, 8, 22–30). The number of aberrations per tumor was 2.5-fold higher in the cell lines than the average number of alterations reported (7.8/case) in the uncultured tumors. However, almost all of the most frequent regions of gain (1q, 3q, 5p, 7p, 7q, 8q,... [rest of the text continues]
17q, 20p, and 20q) and loss (1p, 4p, 8p, 11q, 18p, 18q, Xp, and Xq) identified in our study of the cell lines were also reported to be involved in three or more of the previous studies published on uncultured tumors (Fig. 3). Exceptions were also noted, such as the more frequent involvement of gain of 1p and losses of 10q and 19p in the cell lines.

Many of the cell lines displayed prominent, localized DNA amplifications (Fig. 4). Some of these affected chromosomal regions where...
Fig. 3. An overview of the most common gains and losses reported in our study of 38 breast cancer cell lines as well as 17 published CGH studies of 698 breast tumors. The number of tumors analyzed in each study is indicated in the parenthesis. Green squares, the most common gains in a particular study; red squares, the most common losses. Squares with both colors, a gain and a loss for the same chromosome arm. Approximately 10 of the most common changes were indicated for each study. Despite the similarities of genetic changes observed, many of these studies focused at a specific subtype of breast cancer. Study 1 focused on grade I tumors, studies 2 and 6 focused on grade III tumors, study 3 focused on high-grade ductal carcinoma in situ, study 4 focused on node-positive tumors, studies 5, 7, and 15 focused on node-negative tumors, study 8 focused on hypodiploid cancers, study 12 focused on infiltrating lobular breast cancer, and study 14 focused on metastatic tumors.

Fig. 4. A summary ideogram of regions of high-level amplification by CGH in the 38 breast cancer cell lines. A high-level amplification was defined as small regions (one-to-three chromosomal bands wide) with a highly elevated ratio (ratio >1.4).
genes previously shown to be amplified in breast cancer reside, such as MYC at 8q24 (seven cases), Cyclin-D1/INT2/EMS1 at 11q13 (four cases), ERBB2 at 17q12 (four cases), and FGFR1 at 8p11-p12 (2 cases). Indeed, FISH analysis using specific probes disclosed amplification of MYC in 30%, Cyclin-D1 in 16%, and ERBB2 in 19% of the 37 cell lines. CGH analysis also resulted in the detection of 31 recurrent sites for high-level amplifications of small chromosomal regions not involving the aforementioned established breast cancer oncogenes (Fig. 3). The most common of these regions were 8q23 (37%), 3q25-q26 (24%), 17q22-q24 (16%), 17q23-q24 (16%), 1p13 (11%), 1q32 (11%), 5p13 (11%), 5p14 (11%), 17q12-q21 (11%), and 7q21-q22 (11%).

cDNA Microarray Analyses. The expression patterns of 1236 transcripts were analyzed in four breast cancer cell lines by cDNA microarrays (Fig. 5). A number of highly up-regulated genes were discovered in each cell line: 24 in MCF7 (1.9%), 15 in ZR-75–1-wt (1.2%), 80 in SUM-52 (6.5%), and 52 in SKBR3 (4.2%). The numbers for the down-regulated genes were as follows: 247 in MCF7 (19.9%), 257 in ZR-75–1-wt (20.8%), 174 in SUM-52 (14.1%), and 118 in SKBR3 (9.5%). A list of the top 10 up-regulated genes in each of the four cell lines is shown in Table 1. There were several transcripts that were recurrently up-regulated (in at least two cell lines) and whose chromosomal locations matched with DNA amplification sites seen in the corresponding cell lines. These included MYBL2 at 20q13 (all four cell lines), RCH1 at 17q23 (all four cell lines), TOPO II at 17q21-q22 (all four cell lines), NME1 at 17q21.3 (three cell lines), and CAS at 20q13 (three cell lines). To determine whether the copy numbers of such genes were elevated in the cell lines, FISH analysis was performed using probes for MYBL2, RCH1, TOPO II, and CAS. FISH detected the amplifications of all of these highly up-regulated genes in at least one of the four cell lines analyzed. For example, CAS was amplified in SKBR3 (average of 8 signals/cell) and SUM-52 (14 signals/cell), RCH1 in MCF7 (12 signals/cell) and ZR-75–1-wt (9 signals/cell), TOPO II in SKBR3 (20 signals/cell), and MYBL2 in SKBR3 (13 signals/cell).

DISCUSSION

CGH analysis of 38 breast cancer cell lines identified gains of 1p, 1q, 3q, 5p, 7p, 7q, 8q, 17q, 20p, and 20q and losses of 1p, 4p, 8p, 10q, 11q, 18p, 18q, Xp, and Xq as the most prominent alterations. A comparison between these CGH data from the breast cancer cell lines and published studies in almost 700 uncultured breast cancers (2, 3, 5, 6, 8, 22–30) indicated that many genetic changes in these highly evolved cell lines were similar to those found in the uncultured breast cancers. There were on average 2.5 more alterations/cell line than in the uncultured tumors, but the most common alterations were usually shared between cell lines and uncultured tumors. Similarly, we also identified several recurrent, high-level amplifications in these cell lines, many of which (such as 1q32, 8p11, 8q23, 11q13, 17q23, 17q24, and 20q13) have also been often reported in previous publications of uncultured breast cancers (1). Other amplification sites are novel or may represent changes arising preferentially during the in vitro cell culture, such as 1p13, 7q21, 7q31, 9p23, and 11p13. The set of cell lines analyzed here by CGH provides a resource that should facilitate the cloning of genes involved in such amplifications. The CGH profiles of all of these cell lines were made available on a web site.
DNA amplification is a prominent mechanism of oncogene up-regulation in breast cancer, as best illustrated by the association of \(ERBB2\) amplification and overexpression with poor prognosis in breast cancer (31), a finding that led to the development and implementation of anti-\(ERBB2\) therapy. As illustrated by this and other CGH studies, DNA amplification often takes places at multiple regions of the genome. Identification of all genes involved in these ampiclons would be challenging with traditional techniques. An alternative to positional cloning of amplification target genes is provided by large-scale cDNA microarray analysis of expression levels of numerous genes at once. In our cDNA microarray survey of 1236 genes, we identified several highly overexpressed genes, which mapped to the regions of amplification defined by CGH. These included \(TOPO II\) at 17q22, \(RCH1\) at 17q23, and \(MYBL2\) and \(CAS\) at 20q13. FISH verified the amplifications of all of these highly overexpressed genes in at least one of the four cell lines analyzed. Similarly, we recently identified the ribosomal S6 kinase gene (at 17q23) highly up-regulated in our study of the MCF7 cell line by a cDNA microarray containing 5000 cDNA clones (13). The S6 kinase gene was significantly upregulated and also highly amplified in all of the cell lines with 17q23 amplification.

Finally, although the majority of chromosomal changes identified in the breast cancer cell lines were also present in uncultured tumors, it will be critical to validate the presence of newly identified gene amplifications \textit{in vivo} in uncultured primary breast cancers. Using our recently described high-throughput tissue microarray technology, a series of 372 primary breast cancer specimens were screened for the amplification of the \(MYBL2\) gene, which was identified in this study to be highly overexpressed in breast cancer cell lines. Up to 7% of primary breast cancers studied showed \(MYBL2\) amplification (32). In a similar fashion, we found the S6 kinase gene to be highly amplified in 9% of 668 primary breast tumors. In addition, we were able to show a significant association between amplification of the gene and poor prognosis (13).

The results shown in this study illustrate the powerful approach of defining putative amplification target genes by combining CGH information with results from cDNA microarray analysis, followed by quickly surveying large numbers of uncultured tumors with the tissue microarray technology to study the clinical significance of such newly discovered gene amplifications.

**ACKNOWLEDGMENTS**

We thank Nasser Z. Parsa, Cheryl Ammerman, and Michele Dziubinski for excellent technical assistance and Darryl Leja for help with the technical illustrations. We also thank Vysis, Inc. for the bacterial artificial chromosome illustrations. We also thank Vysis, Inc. for the bacterial artificial chromosome probes.

**REFERENCES**


---

Table 1: Ten most highly up-regulated genes in each cell line

<table>
<thead>
<tr>
<th>Gene name</th>
<th>Chromosome location</th>
<th>Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>ZR-75-1-wt</td>
<td>20q13.1</td>
<td>5.7</td>
</tr>
<tr>
<td>MYBL2*</td>
<td>17q21.3</td>
<td>3.8</td>
</tr>
<tr>
<td>FEN1*</td>
<td>20p13</td>
<td>4.3</td>
</tr>
<tr>
<td>CDC25B</td>
<td>20p13</td>
<td>4.3</td>
</tr>
<tr>
<td>NME1*</td>
<td>17q21.3</td>
<td>3.8</td>
</tr>
<tr>
<td>BRF1*</td>
<td>14q4-q22</td>
<td>3.8</td>
</tr>
<tr>
<td>Topo II*</td>
<td>17q21-q22</td>
<td>3.4</td>
</tr>
<tr>
<td>RCH1</td>
<td>17q23</td>
<td>3.3</td>
</tr>
<tr>
<td>HevGZ*</td>
<td>1p31-p35</td>
<td>3.2</td>
</tr>
<tr>
<td>PAK2*</td>
<td>14q32</td>
<td>3.0</td>
</tr>
<tr>
<td>OCLN</td>
<td>5q12-q13</td>
<td>2.9</td>
</tr>
<tr>
<td>MCF7</td>
<td>5.1</td>
<td></td>
</tr>
</tbody>
</table>

---

* Genes overexpressed in two or more cell lines.


Comparative Genomic Hybridization Analysis of 38 Breast Cancer Cell Lines: A Basis for Interpreting Complementary DNA Microarray Data

Farahnaz Forozan, Eija H. Mahlamäki, Outi Monni, et al.